JOURNAL OF BACTERIOLOGY, Oct. 1995, p. 5680–5685 Vol. 177, No. 19
0021-9193/95/$04.00+0

Rapid Expansion of the Physical and Genetic Map of the Chromosome of *Clostridium perfringens* CPN50

SEI-ICHI KATAYAMA, BRUNO DUPUY, THIERRY GARNIER, AND STEWART T. COLE*

Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 75724 Paris Cédex 15, France

Received 26 April 1995/Accepted 17 July 1995

The physical map of the 3.6-megabase chromosome of *Clostridium perfringens* CPN50 was extended by positioning sites for the endonucleases *Sfi*I and I-*Ceu*I, and in parallel, the gene map was expanded by using a genome scanning strategy. This involved the cloning and sequencing of random chromosomal fragments, identification of the functions of the putative genes by database searches, and then hybridization analysis. The current gene map comprises almost 100 markers, many of which encode housekeeping functions while others are involved in sporulation or pathogenesis. Strikingly, most of the virulence genes were found to be confined to a 1,200-kb segment of the chromosome near *oriC*, while the pleiotropic regulatory locus, *virRS*, was situated toward the putative replication terminus. A comparison of the gene maps of three endospore-forming bacilli, *C. perfringens*, *Clostridium beijerinckii*, and *Bacillus subtilis*, revealed a similar order and distribution of key sporulation and heat shock genes which might reflect an ancient evolutionary relationship.

Clostridium perfringens is a spore-forming, gram-positive anaerobe commonly found in the lower intestinal tracts of humans and other mammals as well as in soil and sewage. *C. perfringens* has been shown to cause a variety of diseases ranging in severity from the frequently fatal gas gangrene to a mild but common form of food poisoning (18, 32). Clinical isolates can be classified into five serotypes, A to E, on the basis of their production of the four lethal typing toxins, the α-, β-, ε-, and ι-toxins (18, 26, 32). In addition to the typing toxins, most strains of *C. perfringens* produce a large variety of other toxins and hydrolytic enzymes, such as perfringolysin O, or θ-toxin, and collagenase, or κ-toxin, (23, 25, 33), that are likely to play a significant role in pathogenesis.

A significant step forward in understanding the molecular genetics of this medically important anaerobe was made some years ago when a physical map of the chromosome of the paradigm strain of *C. perfringens*, CPN50, was constructed with six restriction enzymes and the positions of 28 genes were established (6). Subsequently, the extent of genomic diversity among the various serovars was assessed, three new markers were added, and the variation in the toxin gene repertoire was examined (9).

Although a number of genes have been cloned recently from the clostridia, the rate of identification of new genetic markers has been relatively low. Consequently, in an attempt to rapidly expand the information content of the genome map, we have applied the technique of genome scanning in which random chromosomal fragments are cloned and sequenced and their putative functions are identified by means of database searches. These are then positioned on the physical map by means of hybridization. This strategy was particularly productive and not only led to the identification of 54 new loci but also generated a number of interesting leads for research in pathogenesis because several potential virulence genes were isolated.

* Corresponding author. Mailing address: Unité de Génétique Moléculaire Bactérienne, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cédex 15, France. Phone: (33) 1-45688446. Fax: (33) 1-45688953. Electronic mail address: stcole@pasteur.fr.

MATERIALS AND METHODS

Bacterial strains. *C. perfringens* type A CPN50, also known as BP6K-N5, was used throughout (4, 6). The *Escherichia coli* K-12 strains XL1-blue and JM101 were used for plasmid and M13 preparation, respectively, and standard growth conditions were employed (17).

Genomic DNA preparation and restriction enzyme digests. All genomic DNAs were prepared as described previously (6). *Apa*I, *Avi*II, *Ksp*I, *Nru*I, and *Sma*I were purchased from Boehringer Mannheim, *Mlu*I was purchased from Pharmacia, and *Sfi*I and the intron-encoded endonucleases I-*Ceu*I (24), I-*Til*I (29), and I-*Ppo*I (28) were purchased from New England Biolabs. The DNA in an agarose plug (about 100 μl) was completely digested with 8 to 20 U of most enzymes at 37°C for 3.5 h, while 20 U of *Sfi*I (at 50°C) and *Sma*I (at 25°C) per agarose block was used for 3.5 h to achieve complete digestion.

Electrophoresis and Southern blot analysis. Large restriction fragments were separated by contour-clamped homogeneous electric field gel electrophoresis (10) or field inversion gel electrophoresis as described previously (6). *Saccharomyces cerevisiae* chromosomes (size range, 90 to 1,600 kb) and a mixture of λ concatemers and *Hin*dIII fragments (New England Biolabs) were used as size markers. After electrophoresis, the gels were processed for Southern blotting and the DNA was transferred to Hybond C-extra filters (Amersham) for hybridization analysis as described previously (6, 9).

Genome scanning and probes. To generate new genetic markers for *C. perfringens* CPN50, a shotgun library was constructed in M13 as described recently (14). Chromosomal DNA (50 μg) was sheared by nebulization, end repaired with T4 DNA polymerase (Gibco-BRL) and the Klenow enzyme (Pharmacia), and then fractionated by agarose gel electrophoresis prior to ligation of 0.5- to 2-kb fragments into dephosphorylated *Sma*I-cut M13mp18. Recombinant M13 templates were used in *Taq* polymerase cycled sequencing reactions in a PTC-100 thermal cycler (MJ Research Inc.) with dye-labeled primers (Applied Biosystems) and sequences obtained with an automated DNA sequencer (ABI 373A). An average of 400 nucleotides was obtained from one strand, and most clones were sequenced twice to ensure accuracy. Sequence data were transferred to a SPARC-SUN workstation, and homologous DNA and protein sequences in nonredundant databases at the National Center for Biotechnology Information were identified with the BLAST programs (1, 2). Probes for *C. perfringens recA* and *virR*; *C. perfringens colA*; *C. perfringens nanI*; *C. perfringens sod*; *Clostridium pasteurianum hyd*, *rub*, and *fdx*; and *Clostridium acetobutylicum spo0A* and *spoIVB* were kindly provided by J. Rood, A. Okabe, P. Trieu-Cuot, P. Roggentin, J. Meyer, and M. Young, respectively. Probes for *groEL* and *dnaK* were generated by PCR with primers based on *C. perfringens* sequences (X62914 and X62915 [15]).

RESULTS

Mapping new rare-cutter sites. In an attempt to improve the resolution of the restriction map of the chromosome of *C. perfringens* CPN50, which has sites for *Apa*I, *Avi*II (isoschizomer of *Fsp*I), *Ksp*I (isoschizomer of *Sac*II), *Mlu*I, *Nru*I, and *Sma*I, several new rare-cutting enzymes were tested. Analysis of the *Sfi*I cleavage products of the chromosome of strain

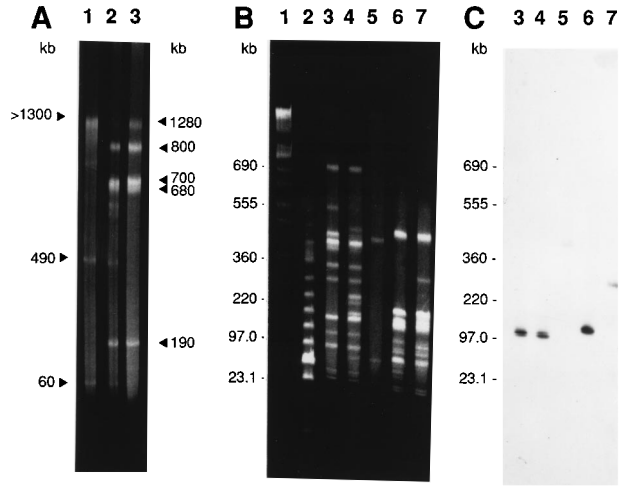

FIG. 1. PFGE analysis and DNA hybridization for the mapping of the *Sfi*I sites of the *C. perfringens* CPN50 genome. (A) Pulsed-field gel of chromosomal DNA from *C. perfringens* CPN50. Lane 1, *Sfi*I digest; lane 2, *Sfi*I plus *Mlu*I double digest; lane 3, *Mlu*I digest. (B) Lane 1, *S. cerevisiae* chromosomes; lane 2, λ concatemers and *Hin*dIII markers; lane 3, *Apa*I digest of *C. perfringens* CPN50 genome; lane 4, *Apa*I plus *Sfi*I double digest; lane 5, *Sfi*I digest; lane 6, *Avi*II plus *Sfi*I double digest; lane 7, *Avi*II digest. (C) The gel from panel B was blotted onto Hybond C membranes and hybridized with a probe for the *eat* gene (Fig. 3). The lanes correspond to those of panel B.

CPN50 by pulsed-field gel electrophoresis (PFGE) suggested that four sites were present as fragments with sizes of >1,300 (two), 490, and 60 kb (Fig. 1A). These sites were initially localized by means of double digestion with *Mlu*I, in which the 1,280- and 700-kb *Mlu*I fragments were cleaved with *Sfi*I (Fig. 1A) and subsequently positioned more precisely by double digestion with the other mapping enzymes. This is illustrated for *Apa*I and *Avi*II (Fig. 1B) by means of hybridization analysis (Fig. 1C).

The chromosomal DNA of *C. perfringens* CPN50 was also digested separately with three intron-encoded endonucleases, I-*Ceu*I (24), I-*Til*I (29), and I-*Ppo*I (28), but only I-*Ceu*I, which generally cuts in rRNA operons (22), reproducibly cleaved DNA to generate a small number of large fragments (Fig. 2).

Mapping I-*Ceu*I sites. *C. perfringens* CPN50 has 10 rRNA (*rrn*) (Fig. 3) operons (16), and I-*Ceu*I generated 10 fragments, A to J, ranging in size from >2 megabases (A) to 9.5 kb (I and J), thus suggesting that the I-*Ceu*I cleavage sites were located in the 23S rRNA genes *rrlA* to *rrlJ*. This was confirmed by the following experiments. *Nru*I cleaves the genome of CPN50 at each *rrn* operon and at a unique site upstream of *rrnJ* (Fig. 3). There was almost no difference between the PFGE profiles obtained after a single digestion with *Nru*I and I-*Ceu*I or after a double digest (Fig. 2A), except in the <10-kb range. The smallest I-*Ceu*I fragment is a 9.5-kb doublet made of fragments I and J, which correspond to DNA from between *rrnB* and *rrnC* and *rrnH* and *rrnJ* (Fig. 2A and 3). By *Nru*I digestion, this was cleaved to yield another doublet with a size of 7.2 kb and 7.4-kb *Nru*I fragments (Fig. 4 gives a schematic interpretation of these results). The 6.5-kb *Nru*I fragment corresponds to the DNA preceding *rrnJ*. These results indicated that all I-*Ceu*I sites were very close to *Nru*I sites and strongly suggested that the sites should be present only in the *rrn* operons.

I-*Ceu*I was shown subsequently to cut once in *rrl* by digesting pBC23, a plasmid carrying a complete *rrn* operon from *C. perfringens* (16), and its cleavage site was precisely localized by sequencing a suitable DNA fragment (accession no. X86518). The faint possibility that another I-*Ceu*I site might exist elsewhere was excluded by appropriate digestions with *Mlu*I, as three of the five *Mlu*I fragments, with sizes of 1,280, 680, and 190 kb, have no *rrn* operon (Fig. 3) and were not cleaved by I-*Ceu*I (Fig. 2B), thereby indicating that the I-*Ceu*I recognition sites were confined to *rrl* genes in *C. perfringens* CPN50, as is the case for several other bacteria (22).

The updated physical map, showing the positions of the *Sfi*I and I-*Ceu*I restriction sites, is shown in Fig. 3.

Mapping cloned clostridial genes by PFGE. Several recently cloned *C. perfringens* genes (*colA*, *dnaK*, *groEL*, *nanI*, *recA*, and *virRS* [15, 23, 25]) were mapped by hybridizing Southern blots of DNA fragments resolved by PFGE after appropriate combinations of single and double digestions with the corresponding probes. Interestingly, the *colA* gene encoding κ-toxin was found to be situated on the same 30-kb *Apa*I-*Sma*I restriction fragment as *nagH*, which codes for another virulence factor, the μ-toxin (8), situated proximal to *oriC*. By contrast, the *virRS* genes encoding the global regulator of several virulence genes were mapped to a region toward the putative replication terminus.

In parallel, probes for known genes from other clostridial species, such as *C. pasteurianum* and *C. acetobutylicum*, were also used at low stringency, although many of these failed to detect the corresponding *C. perfringens* genes, probably as a result of extensive sequence divergence. Heterologous probes enabled the hydrogenase gene, *hyd* (27), and the sporulation genes, *spo0A* and *spoIVB* (5), to be mapped (Fig. 3).

Genome scanning. As the availability of cloned genes was the limiting factor in the analysis of the *C. perfringens* genome, a new strategy, genome scanning, was devised to identify additional genes and accelerate mapping. Short DNA fragments (~1 kb) were cloned randomly from the genome of *C. perfrin-*

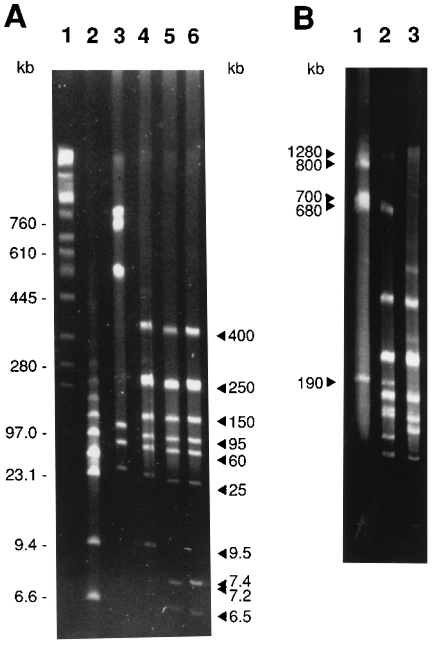

FIG. 2. PFGE analysis and mapping of I-*Ceu*I sites on the *C. perfringens* CPN50 genome. (A) Lane 1, *S. cerevisiae* chromosomes; lane 2, λ concatemers and *Hin*dIII markers; lane 3, I-*Ceu*I digest of *E. coli* W3110 genome; lane 4, I-*Ceu*I digest of *C. perfringens* CPN50 genome; lane 5, I-*Ceu*I plus *Nru*I double digest; lane 6, *Nru*I digest. (B) Pulsed-field gel of chromosomal DNA from *C. perfringens* CPN50. Lane 1, *Mlu*I digest; lane 2, *Mlu*I plus I-*Ceu*I double digest; lane 3, I-*Ceu*I digest.

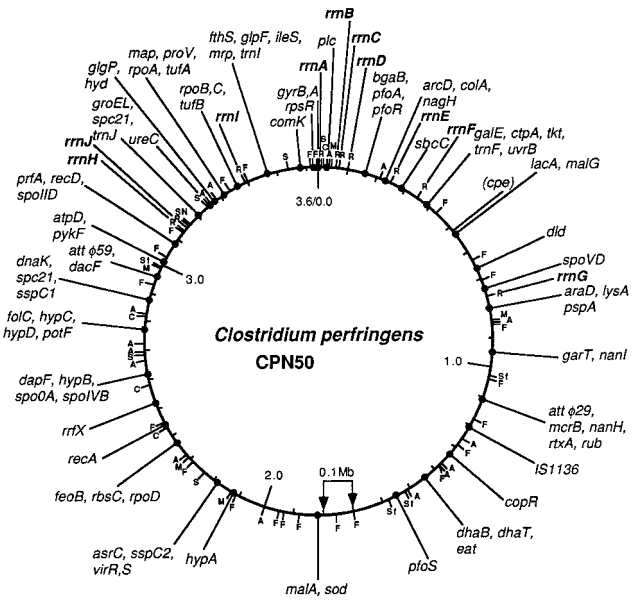

FIG. 3. An updated physical map of the 3.6-megabase (Mb) chromosome of *C. perfringens* CPN50 showing the positions of the various cleavage sites and the *rrn* operons (bold). Restriction enzyme sites: A, *Apa*I; C, *Sac*II (*Ksp*I); F, *Fsp*I (*Avi*II); M, *Mlu*I; N, *Nru*I; R, 2*Sac*II–*Sma*I–*Nru*I–2-I-*Ceu*I–*Sma*I–*Nru*I–I-*Ceu*I; S, *Sma*I; Sf, *Sfi*I. The positions of 97 genetic markers are indicated by the circles situated in the center of the mapping intervals. When several genes map to the same site, their order is arbitrary.

gens CPN50, and their nucleotide sequences were determined (see Materials and Methods). To establish the possible function of these potential genes, homology searches were performed with nonredundant databases on the National Center for Biotechnology Information-BLAST mail server (1). One-third of the random clones (68 of 222) had BLASTX probability scores of $<e^{-12}$ and showed significant sequence similarity to genes or proteins of known function. These new genetic markers and their putative functions are listed in Table 1. Twelve of the clones isolated in this way corresponded to *C. perfringens* genes that had already been described and mapped (*gyrB*, *nagH*, *orfE*, ORF1, *plc* [two], *res*, *rrl* [three], *uviB*, and *colA*). The majority of the functions identified were concerned with housekeeping, although three sporulation genes (*spo0A*, *spoIID*, and *spoVD*) were also isolated (Table 1).

Three of the clones may correspond to genes involved in pathogenesis, as they show strong similarity to known virulence genes from various bacteria (*copR*, *pspA*, and *pfoS*). The product of the *pfoS* gene was highly similar (74% identity in a 146-residue overlap) to that of the perfringolysin O regulatory gene, *pfoR*, from strain NCTC 8237, thus raising the possibility that *pfoS* was allelic to *pfoR* (34). To test this, probes for both genes were used in mapping experiments and were shown to hybridize to two clearly distinct loci (data not shown). Further evidence for the existence of two *pfoR*-like genes was provided by the limited homology at the DNA level (75% in 443 bp).

Expansion of the gene map. Probes were prepared from 54 new clones, and a typical hybridization pattern obtained with one of them (the *eat* gene) is shown in Fig. 1C. In this way, the approximate positions of the corresponding genes were established, and these are indicated on the improved genomic map (Fig. 3). Identical hybridization patterns were obtained when the *spo0A* probes from *C. perfringens* or *C. acetobutylicum* were used, thus underlining the usefulness of heterologous probes based on evolutionarily well-conserved genes (5, 11).

DISCUSSION

The aim of this work was to expand and develop the genome map of *C. perfringens* CPN50, and this was successfully achieved by positioning 14 new sites for endonucleases and establishing the location of 64 new genetic markers in arbitrary map intervals of <100 kb (Fig. 3). The distribution of markers around the chromosome is essentially nonrandom, as there are twice as many mapped genes in the 1.8-megabase segment centered around *oriC* (*gyrA*) compared with those in the region of similar size encompassing the putative replication terminus *terC*. Two possible explanations for this underrepresentation seem likely: either this region of the chromosome is poor in genes or, more probably, there is a bias in the sequences present in current databases, as many of the genes for basic housekeeping functions, such as translation, transcription, and intermediary metabolism, which have been intensively studied in the past, are proximal to *oriC* (12, 21). One means of testing the latter hypothesis would be to use an equivalent number of the shotgun clones obtained in this study whose sequences showed no homology to known genes as probes to see whether the genes are distributed more equally.

Furthermore, as was pointed out previously (6, 9) and extended here, the distribution of genes coding for known toxins, or virulence factors, also appears to be biased. The *plc*, *pfoA*, *colA*, and *nagH* genes encoding the α-, θ-, κ-, and μ-toxins are all situated in a 250-kb stretch close to *oriC*, while the *pspA* gene encoding a putative surface protein homologous to the virulence factor PspA of *Streptococcus pneumoniae* (13) and the *nanH* and *nanI* sialidase genes are ~0.8, ~1, and ~1.2 megabases from *oriC*, respectively (Fig. 3). There is evidence to suggest that sialidase genes are horizontally transferred among bacteria (19), and this, together with the fact that the nearest marker to *nanH* and *nanI* is the attachment site for a lysogenic phage (7), may explain their location on the *C. perfringens* genome map (Fig. 3).

It seems to be a general trend among such gram-positive pathogens as *Listeria monocytogenes*, *Bacillus cereus*, and *C. perfringens* for genes involved in pathogenicity to be located proximal to *oriC* (12). The regulatory locus *virRS* (23, 33), which influences the expression of *plc*, *pfoA*, *colA*, and probably *nanH* and *nanI* as well as unidentified protease and hemagglutinin genes, is situated diametrically opposite the region rich in virulence genes (Fig. 3). This location is consistent with the observation that its products act in *trans* to coordinately

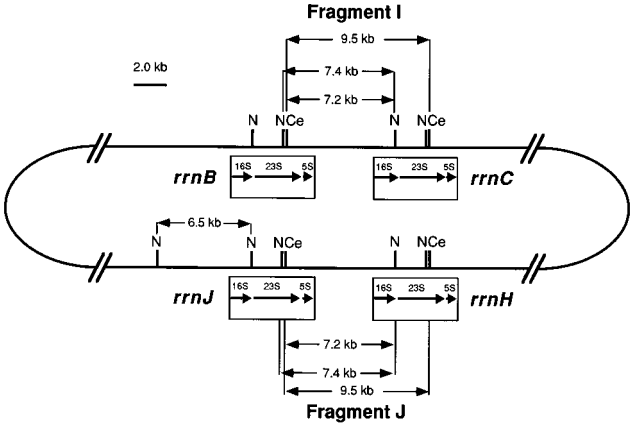

FIG. 4. Schematic interpretation of the mapping data presented in Fig. 2. An abridged version of the chromosome map indicating key restriction sites discussed in the text and 4 of the 10 *rrn* operons is shown.

TABLE 1. Probes of 54 putative genes cloned from *C. perfringens* CPN50

Gene	Clone	Accession no.	Map coordinate(s) (megabases)	Function of similar sequence	Source of similar sequence	Probability score	Accession no. of similar sequence
araD	fcp11	X86527	0.8	L-Ribose-5-phosphate-4-epimerase	*Escherichia coli*	$7.e^{-28}$	P08203
arcD	fcp50	X86520	0.2–0.3	Probable arginine/ornithine antiporter	*Pseudomonas aeruginosa*	$9.e^{-14}$	P18275
aspC	fcp38	X86528	1.1	Aspartate aminotransferase	*Methanobacterium thermoformicicum*	$3.e^{-20}$	JX0306
asrC	fcp5	X86510	2.1–2.2	Anaerobic sulfite reductase subunit	*Salmonella typhimurium*	$1.e^{-48}$	P26476
bgaB	fcp24	X86507	0.1–0.2	β-D-Galactosidase I	*Bacillus circulans*	$6.e^{-43}$	L03425
comK	fcp31	X86485	3.5–3.6	ORFX 5′ of *comK*	*Bacillus subtilis*	$2.e^{-26}$	S70734
copR	fcp23	X86530	1.3–1.4	Transcriptional activator	*Pseudomonas syringae*	$5.e^{-24}$	B47080
ctpA	fcp56	X86497	0.3–0.4	Protease	*Synechocystis* sp.	$7.e^{-18}$	L25250
dacF	fcp48	X86482	2.9–3.0	Penicillin-binding protein	*Bacillus subtilis*	$1.e^{-22}$	P38422
dapF	fcp37	X86511	2.6–2.7	Diaminopimelate epimerase	*Pseudomonas aeruginosa*	$5.e^{-11}$	X78478
dhaB	fcp45	X86532	1.4–1.5	Glycerol dehydratase	*Citrobacter freundii*	$7.e^{-72}$	U09771
dhaT	fcp46	X86487	1.4–1.5	1,3-Propanediol dehydrogenase	*Citrobacter freundii*	$2.e^{-22}$	U09771
dld	fcp40	X86526	0.6–0.7	D-Lactate dehydrogenase	*Lactobacillus plantarum*	$1.e^{-16}$	P26298
feoB	fcp51	X86503	2.3–2.4	Ferrous iron transport protein B	*Escherichia coli*	$2.e^{-16}$	P33650
folC	fcp28	X86524	2.7–2.8	Folylpolyglutamate synthase	*Lactobacillus casei*	$7.e^{-31}$	P15925
fthS	fcp14	X86514	3.4–3.5	Formate-tetrahydrofolate ligase	*Clostridium acidiurici*	$1.e^{-66}$	P13419
galE	fcp10	X86505	0.3–0.4	UDP-glucose-4-epimerase	*Streptococcus thermophilus*	$6.e^{-57}$	P21977
garT	fcp26	X86504	0.9–1	Phosphoribosylglycinamidase-formyl-transferase	*Arabidopsis thaliana*	$5.e^{-17}$	S37105
glgP	fcp22	X86490	3.2–3.3	Glycogen phosphorylase	*Escherichia coli*	$1.e^{-40}$	P13031
glpF	fcp13	X86492	3.4–3.5	Glycerol uptake facilitator protein	*Bacillus subtilis*	$7.e^{-58}$	P18156
gluT-R	fcp41	X86516	1.4–1.5	Glutamate/aspartate transporter	*Homo sapiens*	$7.e^{-26}$	U03504
hypA	fcp35	X86493	2.1–2.2	Hypothetical protein in *nop3-npl3-mts1* 5′ region	*Saccharomyces cerevisiae*	$3.e^{-32}$	P32898
hypB	fcp43	X86496	2.6–2.7	Hypothetical 22.0-kDa protein in *ribT-dacB* intergenic region	*Bacillus subtilis*	$8.e^{-13}$	P35155
hypC	fcp33	X86502	2.7–2.8	Hypothetical 27.5-kDa protein in *dacB-aroC* intergenic region	*Bacillus subtilis*	$8.e^{-24}$	P35163
hypD	fcp34	X86480	2.7–2.8	Hypothetical 51.7-kDa protein in *thrC-dnaK* intergenic region	*Escherichia coli*	$5.e^{-35}$	P30143
ileS	fcp19	X86515	3.4–3.5	Isoleucyl-tRNA synthetase (Ils1)	*Staphylococcus aureus*	$8.e^{-44}$	X75439
IS*1136*	fcp30	X86498	1.2	Insertion element	*Saccharopolyspora erythraea*	$3.e^{-14}$	L07626
lacA	fcp49	X86489	0.5–0.6	Galactose-6-phosphate isomerase	*Streptococcus mutans*	$2.e^{-18}$	P26423
lysA	fcp20	X86512	0.8	Diaminopimelate decarboxylase	*Bacillus methanolicus*	$1.e^{-46}$	L18879
malA	fcp27	X86519	1.8–1.9	Maltose permease	*Bacillus stearothermophilus*	$5.e^{-24}$	L13418
map	fcp2	X86486	3.3–3.4	Methionyl aminopeptidase	*Bacillus subtilis*	$7.e^{-52}$	P19994
mcrB	fcp25	X86483	1.1	5-Methylcytosine-specific restriction	*Escherichia coli*	$7.e^{-16}$	M24927
mrp	fcp32	X86509	3.4–3.5	ATP-binding protein	*Escherichia coli*	$8.e^{-20}$	P21590
pfoS	fcp8	X86525	1.5–1.6	Regulatory protein	*Clostridium perfringens*	$7.e^{-72}$	A43577
potF	fcp39	X86523	2.7–2.8	Periplasmic putrescine-binding protein	*Escherichia coli*	$9.e^{-17}$	P31133
prfA	fcp12	X86479	3.0–3.1	Polypeptide chain release factor	*Salmonella typhimurium*	$4.e^{-51}$	P13654
proV	fcp42	X86517	3.3–3.4	Peripheral membrane protein	*Salmonella typhimurium*	$1.e^{-15}$	P17328
pspA	fcp21	X86522	0.8	Surface protein A	*Streptococcus pneumoniae*	$4.e^{-39}$	A41971
pykF	fcp4	X86495	3.0–3.1	Pyruvate kinase I (PH-1)	*Escherichia coli*	$5.e^{-21}$	P14178
rbsC	fcp3	X86499	2.3–2.4	Membrane ribose-binding protein	*Bacillus subtilis*	$3.e^{-31}$	Z25798
recD	fcp36	X86521	3.0–3.1	Exodeoxyribonuclease V	*Escherichia coli*	$1.e^{-19}$	P04993
rpsR	fcp6	X86513	3.6–0	Ribosomal protein S18 (RS18)	*Bacillus stearothermophilus*	$4.e^{-28}$	P10806
rrfX	fcp16	X86481	2.5–2.6	Ribosome releasing factor (RRF)	*Escherichia coli*	$3.e^{-37}$	P16174
rub	fcp17	X86500	1.1	Rubredoxin	*Clostridium perfringens*	$1.e^{-28}$	P14072
sbcC	fcp29	X86484	0.2–0.3	Exonuclease	*Escherichia coli*	$8.e^{-14}$	A43750
sod	fcp54	Z49062	1.8–1.9	Superoxide dismutase	*Bacteroides fragilis*	$3.e^{-29}$	D13756
spc21	fcp53	X86488	2.8–2.9	Microsomal signal peptidase	*Canis familiaris*	$9.e^{-11}$	P13679
spo0A	fcp44	X86491	2.6–2.7	Phosphorylation-activated transcription factor	*Bacillus subtilis*	$3.e^{-31}$	U09979
spoIID	fcp9	X86529	3.0–3.1	Stage II sporulation protein D	*Clostridium acetobutylicum*	$1.e^{-32}$	S27530
spoVD	fcp18	X86501	0.7–0.8	Penicillin-binding protein	*Bacillus subtilis*	$3.e^{-40}$	Z25865
tkt	fcp7	X86506	0.3–0.4	Transketolase (TK)	*Homo sapiens*	$6.e^{-39}$	P29401
trnF	fcp47	X86508	0.4	Proline tRNA	*Bacillus subtilis*	0.00077	M27310
trnI	fcp47	X86508	3.2	tRNA cluster	*Bacillus subtilis*	0.00077	M27310
trnJ	fcp47	X86508	3.4–3.5	tRNA cluster	*Bacillus subtilis*	0.00077	M27310
ureC	fcp1	X86494	3.2–3.3	Probable phosphomannomutase	*Mycobacterium leprae*	$7.e^{-43}$	U00020
uvrB	fcp55	X86531	0.3–0.4	DINA protein	*Bacillus subtilis*	$1.e^{-49}$	P37954

regulate this group of genes. This is unlike the situation with *pfoR*, which encodes a *cis*-acting regulator of the adjacent gene, *pfoA* (Fig. 3) (34). The isolation of *pfoS*, a gene homologous to *pfoR*, provides a promising lead for further research, as it is conceivable that the coding sequence for a potential virulence factor, such as the θ-toxin, may be linked. This possibility is currently being explored.

As detailed physical and gene maps of the chromosomes of

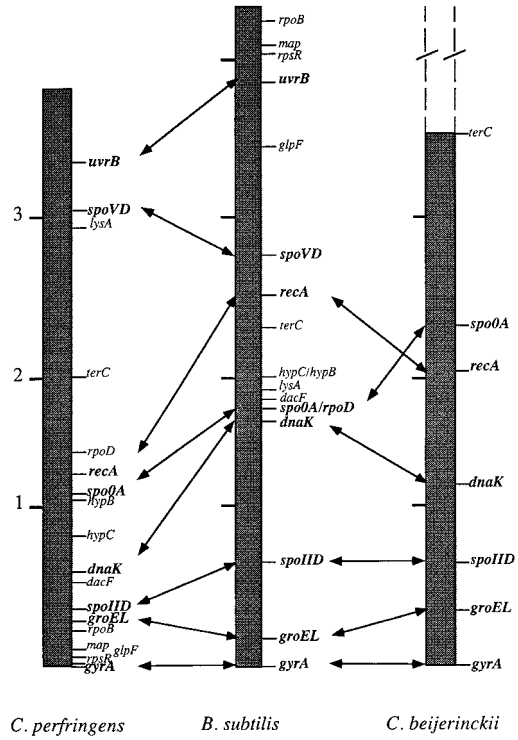

FIG. 5. Comparative analysis of maps of the chromosomes of the endospore-forming bacilli *C. perfringens*, *B. subtilis*, and *C. beijerinckii* drawn from this study and other sources (3, 6, 20, 36). The maps are aligned about the *gyrA* gene, a marker for *oriC*, and the positions of highly conserved single copy genes are shown, with key genes present in more than one species linked by doubleheaded arrows. Note that the position of *terC* has not been physically mapped in *C. perfringens* and *C. beijerinckii* but is inferred as being roughly diametrically opposite *oriC*. In all cases, although *spo0A* and *spoIVB* are linked (5), only *spo0A* is shown. For reasons of space, only half of the chromosome of the 6.7-megabase chromosome of *C. beijerinckii* is shown (36). The numbers at the left indicate megabases.

three endospore-forming bacteria, *Bacillus subtilis* (3, 20), *C. perfringens* (Fig. 3), and *C. beijerinckii* (36), are now available, meaningful comparisons of the distributions of sporulation genes were possible. These were especially informative when the maps of *B. subtilis* and *C. perfringens* were compared, because despite many differences, an apparent relationship was detected in the localization of four sporulation genes, *spoIID*, *spo0A-spoIVB*, and *spoVD*, and the heat shock genes *groELS* and *dnaK* (Fig. 5). In both cases, the relative positions and orders of these markers were similar. Furthermore, the *recA* and *rpoD* genes are also centrally located on both maps. When this analysis was extended to include *C. beijerinckii* (36), essentially similar findings were made for this set of genes, suggesting that this may be a general trend for members of the family *Bacillaceae*. As more genome maps become available, it will be interesting to see whether this feature, which may indicate an ancient evolutionary relationship, holds true. A comparative analysis of the genetic maps of six species of the genus *Bacillus* has been undertaken and has revealed global homologies like those reported here but also many local differences in gene distribution and order (35). Interestingly, the positions of the *ssp* genes, which are expressed during sporulation, were found to be conserved (35).

In conclusion, the genome scanning approach used in this study has proved to be a rapid and profitable means of expanding the gene map of *C. perfringens* and providing new information about the metabolic and pathogenic potential of this anaerobic bacterium. Similar projects in which large numbers of random clones were sequenced have been undertaken with *Mycoplasma genitalium* and have also yielded vast amounts of valuable information about the physiology and biochemistry of this organism (30). Very recently (31), these sequence-tagged sites have also been positioned on the genomic map, thus indicating the general utility of the genome scanning approach.

ACKNOWLEDGMENTS

We are very grateful to Akinobu Okabe, Jacques Meyer, Peter Roggentin, Julian Rood, Patrick Trieu-Cuot, and Mike Young for kind gifts of probes and helpful discussions.

This investigation received financial support from the Japanese Ministry of Education, the Legs van Straelen, and the Institut Pasteur.

REFERENCES

1. **Altschul, S., W. Gish, W. Miller, E. Myers, and D. Lipman.** 1990. A basic local alignment search tool. J. Mol. Biol. **215:**403–410.
2. **Altschul, S. F., M. S. Boguski, W. Gish, and J. C. Wooton.** 1994. Issues in searching molecular sequence databases. Nat. Genet. **6:**119–129.
3. **Anagnostopoulos, C., P. J. Piggot, and J. A. Hoch.** 1993. The genetic map of *Bacillus subtilis*, p. 425–461. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, D.C.
4. **Bréfort, G., M. Magot, H. Ionesco, and M. Sebald.** 1977. Characterization and transferability of *Clostridium perfringens* plasmids. Plasmid **1:**52–66.
5. **Brown, D. P., L. Ganova-Raeva, B. D. Green, S. R. Wilkinson, M. Young, and P. Youngman.** 1994. Characterization of *spo0A* homologues in diverse *Bacillus* and *Clostridium* species identifies a probable DNA-binding domain. Mol. Microbiol. **14:**411–426.
6. **Canard, B., and S. T. Cole.** 1989. Genome organization of the anaerobic pathogen *Clostridium perfringens*. Proc. Natl. Acad. Sci. USA **86:**6676–6680.
7. **Canard, B., and S. T. Cole.** 1990. Lysogenic phages of *Clostridium perfringens*: mapping of the chromosomal attachment sites. FEMS Microbiol. Lett. **66:**323–326.
8. **Canard, B., T. Garnier, B. Saint-Joanis, and S. T. Cole.** 1994. Molecular genetic analysis of the *nagH* gene encoding a hyaluronidase of *Clostridium perfringens*. Mol. Gen. Genet. **243:**215–224.
9. **Canard, B., B. Saint-Joanis, and S. T. Cole.** 1992. Genomic diversity and organisation of virulence genes in the pathogenic anaerobe *Clostridium perfringens*. Mol. Microbiol. **6:**1421–1429.
10. **Chu, G., D. Vollrath, and R. W. Davis.** 1986. Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science **234:** 1582–1585.
11. **Cole, S. T., and B. Canard.** 1993. Genome mapping of *Clostridium perfringens* type A strains, p. 255–263. *In* M. Sebald (ed.), Genetics and molecular biology of anaerobes. Springer-Verlag, New York.
12. **Cole, S. T., and I. Saint-Girons.** 1994. Bacterial genomics. FEMS Microbiol. Rev. **14:**139–160.
13. **Crain, M. J., W. D. Waltman II, J. S. Turner, J. Yother, D. K. Talkington, L. S. McDaniel, B. M. Gray, and D. E. Briles.** 1990. Pneumococcal surface protein A (PspA) is serologically highly variable and is expressed by all clinically important capsular serotypes of *Streptococcus pneumoniae*. Infect. Immun. **58:**3293–3299.
14. **Fsihi, H., and S. T. Cole.** The *Mycobacterium leprae* genome: systematic sequence analysis identifies key catabolic enzymes, ATP-dependent transport systems and a novel *polA* locus associated with genomic variability. Mol. Microbiol., in press.
15. **Galley, K. A., B. Singh, and R. S. Gupta.** 1992. Cloning of HSP70 (*dnaK*) gene from *Clostridium perfringens* using a general polymerase chain reaction based approach. Biochim. Biophys. Acta **1130:**203–208.
16. **Garnier, T., B. Canard, and S. T. Cole.** 1991. Cloning, mapping, and molecular characterization of the rRNA operons of *Clostridium perfringens*. J. Bacteriol. **173:**5431–5438.
17. **Garnier, T., and S. T. Cole.** 1986. Characterization of a bacteriocinogenic plasmid from *Clostridium perfringens* and molecular genetic analysis of the bacteriocin-encoding gene. J. Bacteriol. **168:**1189–1196.
18. **Hatheway, C. L.** 1990. Toxigenic clostridia. Clin. Microbiol. Rev. **3:**66–98.
19. **Hoyer, L. L., A. C. Hamilton, S. M. Steenbergen, and E. R. Vimr.** 1992. Cloning, sequencing and distribution of the *Salmonella typhimurium* LT2 sialidase gene, *nanH*, provides evidence for interspecies gene transfer. Mol. Microbiol. **6:**873–884.
20. **Itaya, M.** 1993. Physical map of the *Bacillus subtilis* 168 chromosome, p. 463–471. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, D.C.

21. **Krawiec, S., and M. Riley.** 1990. Organization of the bacterial chromosome. Microbiol. Rev. **54:**502–539.
22. **Liu, S.-L., A. Hessel, and K. E. Sanderson.** 1993. Genomic mapping with I-*Ceu*I, an intron-encoded endonuclease specific for genes for ribosomal RNA, in *Salmonella* spp., *Escherichia coli*, and other bacteria. Proc. Natl. Acad. Sci. USA **90:**6874–6878.
23. **Lyristis, M., A. E. Bryant, J. Sloan, M. M. Awad, I. T. Nisbet, D. L. Stevens, and J. I. Rood.** 1994. Identification and molecular analysis of a locus that regulates extracellular toxin production in *Clostridium perfringens*. Mol. Microbiol. **12:**761–777.
24. **Marshall, P., and C. Lemieux.** 1991. Cleavage pattern of the homing endonuclease encoded by the fifth intron in the chloroplast large subunit rRNA-encoding gene of *Chlamydomonas eugametos*. Gene **104:**241–245.
25. **Matsushita, O., K. Yoshihara, S. Katayama, J. Minami, and A. Okabe.** 1994. Purification and characterization of a *Clostridium perfringens* 120-kilodalton collagenase and nucleotide sequence of the corresponding gene. J. Bacteriol. **176:**149–156.
26. **McDonel, J. L.** 1986. Toxins of Clostridium perfringens types A, B, C, D and E, p. 477–517. *In* F. Dorner and J. Drews (ed.), Pharmacology of bacterial toxins. Pergamon Press, Oxford.
27. **Meyer, J., and J. Gagnon.** 1991. Primary structure of hydrogenase I from *Clostridium pasteurianum*. Biochemistry **30:**9696–9704.
28. **Muscarella, D. E., E. L. Ellison, B. M. Ruoff, and V. M. Vogt.** 1990. Characterization of I-*Ppo*, an intron-encoded endonuclease that mediates homing of a group I intron in the ribosomal DNA of *Physarum polycephalum*. Mol. Cell. Biol. **10:**3386–3396.
29. **Perler, F. B., D. G. Comb, W. E. Jack, L. S. Moran, B. Qiang, R. B. Kucera, J. Benner, B. E. Slatko, D. O. Nwankwo, S. K. Hempstead, C. K. S. Carlow, and H. Jannasch.** 1992. Intervening sequences in an Archaea DNA polymerase gene. Proc. Natl. Acad. Sci. USA **89:**5577–5581.
30. **Peterson, S. N., P.-C. Hu, K. F. Bott, and C. A. Hutchison III.** 1993. A survey of the *Mycoplasma genitalium* genome by using random sequencing. J. Bacteriol. **175:**7918–7930.
31. **Peterson, S. N., T. Lucier, K. Heitzman, E. A. Smith, K. F. Bott, P.-C. Hu, and C. A. Hutchison III.** 1995. Genetic map of the *Mycoplasma genitalium* chromosome. J. Bacteriol. **177:**3199–3204.
32. **Rood, J. I., and S. T. Cole.** 1991. Molecular genetics and pathogenesis of *Clostridium perfringens*. Microbiol. Rev. **55:**621–648.
33. **Shimizu, T., W. Ba-Thein, M. Tamaki, and H. Hayashi.** 1994. The *virR* gene, a member of a class of two-component response regulators, regulates the production of perfringolysin O, collagenase, and hemagglutinin in *Clostridium perfringens*. J. Bacteriol. **176:**1616–1623.
34. **Shimizu, T., A. Okabe, J. Minami, and H. Hayashi.** 1991. An upstream regulatory sequence stimulates expression of the perfringolysin gene of *Clostridium perfringens*. Infect. Immun. **59:**137–142.
35. **Vary, P. S.** 1993. The genetic map of *Bacillus megaterium*, p. 475–481. *In* A. L. Sonenshein, J. A. Hoch, and R. Losick (ed.), *Bacillus subtilis* and other gram-positive bacteria. American Society for Microbiology, Washington, D.C.
36. **Wilkinson, S. R., and M. Young.** 1995. Physical map of the *Clostridium beijerinckii* (formerly *Clostridium acetobutylicum*) NCIMB 8052 chromosome. J. Bacteriol. **177:**439–448.